



Evolution of a Dominant Natural Isolate of *Escherichia coli* in the Human Gut over the Course of a Year Suggests a Neutral Evolution with Reduced Effective Population Size

Mohamed Ghalayini,^{a,b,c} Adrien Launay,^{a,d} Antoine Bridier-Nahmias,^{a,d} Olivier Clermont,^{a,d} Erick Denamur,^{a,d,e} Mathilde Lescat,^{a,b,c} Olivier Tenaillon^{a,d}

^aINSERM, IAME, UMR 1137, Paris, France

^bUniversité Paris Nord, Sorbonne Paris Cité, Bobigny, France

^cAP-HP, Hôpitaux Universitaires Paris Seine Saint-Denis, Bondy, France

^dUniversité Paris Diderot, Sorbonne Paris Cité, Paris, France

^eAP-HP, Hôpitaux Universitaires Paris Nord Val de Seine, Paris, France

ABSTRACT *In vitro* and *in vivo* evolution experiments on *Escherichia coli* revealed several principles of bacterial adaptation. However, few data are available in the literature describing the behavior of *E. coli* in its natural environment. We attempted here to study the evolution in the human gut of a commensal dominant *E. coli* clone, ED1a belonging to the B2 phylogroup, through a longitudinal genomic study. We sequenced 24 isolates sampled at three different time points within a healthy individual over almost a year. We computed a mutation rate of 6.90×10^{-7} mutations per base per year of the chromosome for *E. coli* ED1a in healthy human gut. We observed very limited genomic diversity and could not detect any evidence of selection, in contrast to what is observed in experimental evolution over a similar length of time. We therefore suggest that ED1a, being well adapted to the healthy human gut, evolves mostly neutrally with a low effective population size (N_e of ≈ 500 to 1,700).

IMPORTANCE In this study, we follow the genomic fate of a dominant clone of *Escherichia coli* in the human gut of a healthy individual over about a year. We could compute a low annual mutation rate that supports low diversity, and we could not retrieve any clear signature of selection. These observations support a neutral evolution of *E. coli* in the human gut, compatible with a very limited effective population size that deviates drastically with the observations made previously in experimental evolution.

KEYWORDS ED1a, *Escherichia coli*, human gut, molecular evolution, mutation rate, neutral evolution, replication rate

Resistance to antibiotics has sadly revealed that microbial evolution is an active process that can be witnessed in the short term (1–4). Antibiotic pressures represent an extreme form of selection. As a result, strong and fast selective responses are expected and documented. Much less is known about the pace of microbial evolution in the wild under milder selective pressures, even for bacterial species of medical interest such as *Escherichia coli*, a commensal of the gut and a versatile pathogen.

In vitro and *in vivo* evolution experiments on *E. coli* have revealed several principles of bacterial adaptation: (i) selection regularly favors an increased mutation rate that resulted in an improved adaptation rate (5–7), (ii) global regulators with pleiotropic effects are often recruited in the first stages of adaptation (8–11), (iii) lineages evolved under similar conditions evolve frequently through modification of the same genes albeit with different mutations (12–14), (iv) traces of selection overwhelm the genomic

Received 29 October 2017 Accepted 22 December 2017

Accepted manuscript posted online 5 January 2018

Citation Ghalayini M, Launay A, Bridier-Nahmias A, Clermont O, Denamur E, Lescat M, Tenaillon O. 2018. Evolution of a dominant natural isolate of *Escherichia coli* in the human gut over the course of a year suggests a neutral evolution with reduced effective population size. *Appl Environ Microbiol* 84:e02377-17. <https://doi.org/10.1128/AEM.02377-17>.

Editor Marie A. Elliot, McMaster University

Copyright © 2018 American Society for Microbiology. All Rights Reserved.

Address correspondence to Mohamed Ghalayini, mohamed.ghalayini@inserm.fr.

M.G. and A.L. contributed equally to this work.

changes observed (6), and, finally, (v) the importance of selection is linked to maladaptation of the strain used in the experiment (14). Though globally consistent across experiments, these observations emerged nevertheless under artificial conditions. Even in the case of evolution in the mouse gut (11, 14), an environment much closer to *E. coli*'s habitat than test tubes, the caged life of the animals, the consistency of their diet, or the use of streptomycin to maintain *E. coli* are factors that are driving the process away from natural conditions. Hence, the relevance of these principles of adaptation under natural conditions, especially under human commensal conditions, remains questionable.

E. coli is a versatile species, known as both a widespread gut commensal of the vertebrates and a dangerous pathogen that can also be retrieved in the environment (15). As a pathogen *E. coli* is responsible for about 1 million human deaths yearly due to intraintestinal and now mostly extraintestinal diseases (16). Among the seven phylogenetic groups, A, B1, B2, C, D, E, and F (17), that compose the species, B2 and D are the ones most often associated with extraintestinal diseases. The great majority of *E. coli* strains belonging to genetic group B2 are highly virulent in a mouse model of extraintestinal infection (18). Yet one subgroup (subgroup VIII, corresponding to sequence type 452 [ST452] or ST149 using the Achtman or Pasteur Institute scheme, respectively, [19], and exhibiting the O81:H27 serotype) within this group does not exhibit any extraintestinal virulence in a mouse model of sepsis (20). Moreover, this subgroup, which includes archetypal clone *E. coli* ED1a, happens to be specific to the human gut (21). When present, strains from that nonpathogenic B2 group are dominant compared to other strains of the phylogenetic group B2 and persist for long periods in the human gut. This suggests that these strains perform well in the human gut (21).

To circumvent the limitation of the *in vitro* and *in vivo* studies of experimental evolution, we attempt here to study the evolution of *E. coli* in its natural environment, the human gut, through a longitudinal genomic study. This approach was previously used in a 3-year-long study of the evolution and transmission of an *E. coli* clone in a household of six members, including a dog. The investigators followed a B2 clone (named clone D) that was not dominant but that could be recovered 14 times at different time points. They showed a high transmission rate of the clone within the household, a low mutation rate, limited diversity, and no clear traces of adaptation (22). In the present study, we focus on the evolution of a dominant clone retrieved over a 1-year period within a healthy individual.

To study the evolution and adaptation in the human gut of the commensal clone ED1a, we sequenced 24 isolates sampled at three different time points within an individual. The aim of our retrospective study was to identify the evolutionary forces shaping the evolution of its genome in the human gut over almost a year.

RESULTS

Dominance and persistence of *E. coli* ED1a. Eight isolates of *E. coli* were randomly selected after plating the feces on Drigalski medium at three time points: day 0 (D0), day 211 (D211), and day 315 (D315). All isolates belonged to the phylogenetic group B2, subgroup VIII, and exhibited the O81 type, confirming their belonging to the ED1a clone (21). The lack of alternative *E. coli* genotypes suggests that the *E. coli* niche was dominated, if not fully invaded, by the ED1a clone. This also suggests a persistence of this dominance over almost 1 year in healthy human gut although the number of clones sampled per time point ($n = 8$) was relatively small.

Description of *E. coli* ED1a evolution. *E. coli* ED1a, in addition to its chromosomal genome (5,209,548 bp), possesses a conjugative plasmid, pECOED, of 119,594 bp (23). On the 24 isolate genomes analyzed by the breseq pipeline, we identified 17 different mutations, among which 2 were located on the plasmid. We observed eight mutations among the eight isolates from D0 (including two deletions, four nonsynonymous single nucleotide polymorphisms [SNPs], and two synonymous SNPs), five mutations among the eight isolates from D211 (including four nonsynonymous SNPs and one SNP giving

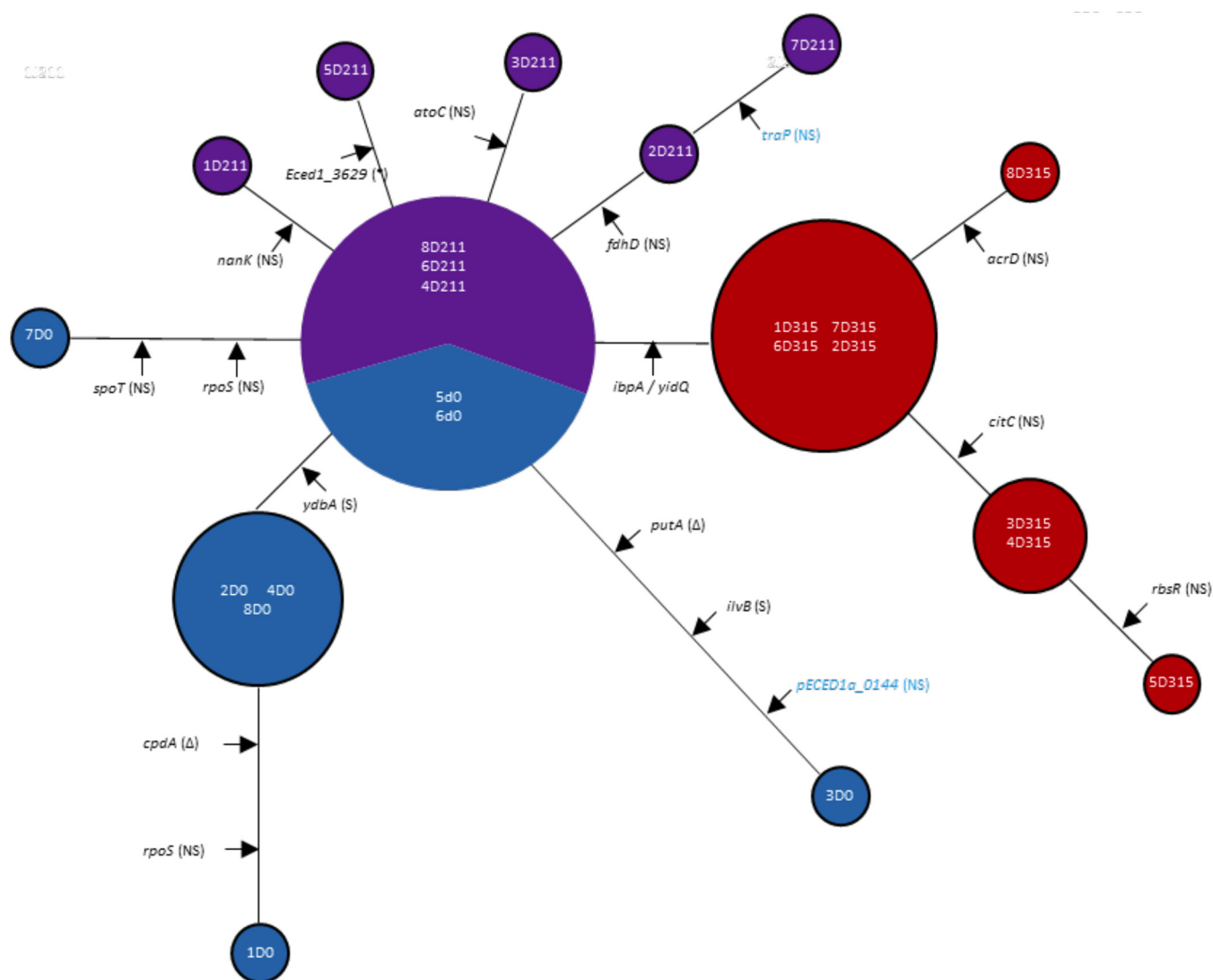


FIG 1 Genomic diversity of *E. coli* ED1a over almost a year in a healthy human gut. We represented a maximum likelihood unrooted tree relating all isolates sampled over a year reconstructed with the packages *ape* (55) and *phangorn* (56) in R software (34). The length of the branches is proportional to the number of mutations separating two samples. The diameters of circles are proportional to the numbers of isolates that are identical. Color codes are as follows: blue, isolates sampled at day 0 (D0); violet, isolates sampled at day 211 (D211); red, isolates sampled at day 315 (D315). The intergenic mutations are represented by two genes separated by a slash. The nonsynonymous and synonymous SNPs are marked by (NS) and (S), respectively. Deletion mutations (12 bp in the *putA* gene and 3 bp in a repeated region of the *cpdA* gene) are indicated with a Greek delta (Δ). A stop codon is indicated with an asterisk (*). The mutations occurring on the plasmid are shown in blue.

a stop codon), and four mutations (including three nonsynonymous SNPs and one SNP in intergenic region) among the eight isolates from D315. Several isolates had identical genomes (Fig. 1). While 13 out of the 17 mutations were specific to one isolate (singletons), four mutations were shared by several isolates. These mutations included an intergenic mutation between *ibpA* and *yidQ* that seems to have been fixed between D211 and D315 as all D315 isolates showed this mutation while none had it at D211 (Table 1).

In sum, SNPs represented 15 of the 17 mutations observed (88% of the mutations), with 11 nonsynonymous mutations, 2 synonymous mutations, 1 stop codon mutation, and 1 intergenic mutation. Among the SNPs, there were seven transitions and eight transversions. The two non-SNP mutations were deletions found at D0: a deletion of 3 bp within a repeat region of *cpdA* in isolate 1 obtained on D0 (1D0) and a 12-bp deletion in *putA* in sample 3D0. No large deletion, large insertion, or transposon insertions were detected (Table 1).

TABLE 1 Summary of mutations that were found in the 24 isolates of *E. coli* ED1a sampled over almost a year in a healthy human gut

Gene ^a	Position (nt) ^b	Mutation	Mutation effect	Type of mutation (no. of instances)	Isolate no.	Gene product(s)
<i>acrD</i> →	2852335	A → C	D164A (GAC → GCC)	Nonsynonymous SNP (1)	8D315	Aminoglycoside/multidrug efflux system
<i>atoC</i> →	2612636	T → G	I129S (ATC → AGC)	Nonsynonymous SNP (1)	3D211	Fused response regulator of <i>ato</i> operon in a two-component system with AtoS; response regulator; sigma54 interaction protein
<i>citC</i> ←	637957	T → G	T350P (ACC → CCC)	Nonsynonymous SNP (3)	3D315, 4D315, 5D315	Citrate lyase synthetase
<i>cpdA</i> ←	3623372	(GCA)4 → 3	Deletion of A	In-phase deletion (1)	1D0	3'-5' Cyclic AMP phosphodiesterase
<i>ECED1_3629</i> ←	3559530	C → A	E146* (GAG → TAG) ^c	Stop codon SNP (1)	5D211	Conserved hypothetical protein; putative membrane protein
<i>fdhD</i> →	4541860	C → T	A114V (GCG → GTG)	Nonsynonymous SNP (2)	2D211, 7D211	Formate dehydrogenase formation protein
<i>ibpA</i> ←/→ <i>yidQ</i>	4305206	C → T	NA	Intergenic SNP (8)	1D315, 2D315, 3D315, 4D315, 5D315, 6D315, 7D315, 8D315	Heat shock chaperone/conserved hypothetical protein; putative outer membrane protein
<i>ilvB</i> ←	4294479	C → A	A88A (GCG → GCT)	Synonymous SNP (1)	3D0	Acetolactate synthase I large subunit
<i>nanK</i> ←	3804472	T → C	T32A (ACG → GCG)	Nonsynonymous SNP (1)	1D211	Putative N-acetylmannosamine kinase
<i>putA</i> ←	1173242	Δ12 bp	Deletion of EEWQ	In-phase deletion (1)	3D0	Fused DNA-binding transcriptional regulator; proline dehydrogenase; pyrroline-5-carboxylate dehydrogenase
<i>rbtR</i> →	4375084	T → G	V18G (GTT → GGT)	Nonsynonymous SNP (1)	5D315	DNA-binding transcriptional repressor of ribose metabolism
<i>rpoS</i> ←	3132176	A → C	I198S (ATC → AGC)	Nonsynonymous SNP (1)	1D0	RNA polymerase, sigma S (sigma 38) factor
<i>rpoS</i> ←	3132465	C → T	V102 M (GTG → ATG)	Nonsynonymous SNP (1)	7D0	RNA polymerase, sigma S (sigma 38) factor
<i>spoT</i> →	4255773	G → A	E98K (GAG → AAG)	Nonsynonymous SNP (1)	7D0	Bifunctional (p)ppGpp synthetase II and guanosine-3',5'-bis pyrophosphate 3'-pyrophosphohydrolase
<i>ydbA</i> →	1543455	G → T	G413G (GGG → GGT)	Synonymous SNP (4)	1D0, 2D0, 4D0, 8D0	Putative autotransported outer membrane protein involved in cell adhesion
<i>pECED1a_0115</i> ←	85479	C → T	D115N (GAT → AAT)	Nonsynonymous SNP (1)	7D211	Conserved hypothetical protein
<i>pECED1a_0144</i> →	112511	T → C	S49P (TCC → CCC)	Nonsynonymous SNP (1)	3D0	Hypothetical protein; putative membrane protein

^aArrows indicate the orientations of the genes; a slash indicates the intergenic region. The mutations present on the plasmid are preceded by the prefix *pECED1a*. *pECED1a_0115* corresponds to *traP*.^bnt, nucleotide.^cThe asterisk (*) indicates a stop codon.

Analyses of prophage and supplementary plasmid sequences in *E. coli* ED1a.

We observed several reads in all isolates that did not map to the ED1a genome reference. We searched for mobile elements that could explain these unmapped reads.

In the reads that did not map on the genome, the only phage we could detect was PhiX174. Given that Illumina recommends use of PhiX174 as a sequencing control in proportions varying between 1% and 40%, these reads are likely to result from a contamination of our fastq files by PhiX reads. This phenomenon is already documented in the literature and seems to be pervasive among sequenced genomes deposited in the Genome Online Database (<https://gold.jgi.doe.gov/>) (24).

To detect some plasmid DNA, we used plasmidSPAdes on the contigs assembled from the reads that did not map on the ED1a chromosome and large plasmid. The plasmidSPAdes algorithm could not reconstruct any full circular plasmid but, after the sequences were annotated with Prokka, the genes we identified were found to be common among mobilizable plasmids (25). The absence of a good reference for these potentially plasmidic sequences made it impossible to use breseq to detect genetic variations between the isolates. To overcome this hurdle, we used a reference-free SNP detection tool called discoSnp++ and did not detect any polymorphic locus (26).

We performed a PCR assay with primers specific to the hypothetical plasmid matching an ORF coding a protein belonging to the *mobA-mobL* superfamily and obtained an amplicon in isolates 6D0, 6D211, and 6D315 showing the presence of this hypothetical plasmid from the first isolation of the strain to the end of the experiment almost 1 year later. Thus, we observed a small additional plasmid with no mutation.

***E. coli* ED1a mutation rate.** Mutation rates can be estimated simply by assuming that isolates evolve independently. To do so, we used the different mutations we observed in D211 isolates (four mutations, or five mutations including the plasmid) and in D315 isolates (four mutations), considering D0 and D211 isolates, respectively, to be their ancestors. We hypothesized consequently that mutations at D211 occurred during the interval between D0 and D211 and that mutations at D315 occurred during the interval between D211 and D315. We computed an average chromosomal mutation rate of 2.51×10^{-7} mutations per base per year, assuming that we covered the whole chromosomal genome, and a rate of 2.66×10^{-7} mutations per base per year taking into account the plasmid (Table 2).

This rough estimate of mutation rate that is commonly used neglects, however, that isolates are not independent but connected through a genealogy. Such genealogy can be inferred assuming the population is homogenous and can consequently be modeled according to coalescent theory. As our isolates are sampled from a single host, this hypothesis seems relevant. We therefore computed the mutation rate using a Bayesian approach implemented in BEAST (27). A mutation rate of 6.90×10^{-7} mutations per base per year (95% confidence interval [CI], 3.14×10^{-7} to 1.40×10^{-6} mutations per base per year) was obtained on the chromosomal genome, and a rate of 7.18×10^{-7} mutations per base per year (95% CI, 3.28×10^{-7} to 1.44×10^{-6} mutations per base per year) was obtained if the plasmid was included (Table 2). These estimates differ significantly from the estimates that rely on the independence of mutations, with the latter being outside the 95% confidence interval of the former.

We therefore applied the BEAST approach to another data set in which the authors followed diversification of a clone within a household over 3 years (22). They used a method that partially took into account the genealogy to compute a rate of 2.26×10^{-7} mutations per base per year. However, by calculating the mutation rate based on the coalescent theory, we obtained a mutation rate of 1.40×10^{-7} mutations per base per year (95% CI, 2.97×10^{-8} to 2.77×10^{-7} mutations per base per year), assuming a 5,038,386-bp-length clone D genome (Table 2). These two mutation rate estimations are quite close, but the Bayesian approach seems to be more suitable and accurate; indeed, there remained an uncertainty as to the period of time at which the mutations appeared.

Once the mutation rate per base per year (μ_{est}) was computed, we estimated the rate of replication (G) of *E. coli* ED1a in healthy human gut using the formula

TABLE 2 Comparison of mutation rates per base per year and numbers of generations per day estimated by different models of evolution on different data sets from evolution studies^a

Strain	Phylogenetic group	Independent evolution of lineage (no. of mutations/base/yr [95% CI])	Coalescent theory of evolution			Reference
			No. of mutations/base/yr (95% CI)	No. of generations/day (95% CI) ^c		
<i>E. coli</i> ED1a (chromosome only)	B2	2.51×10^{-7}	6.90×10^{-7} (3.14×10^{-7} – 1.40×10^{-6})	6.06 (2.88–12.65)–21.24 (9.66–43.28)	This study	
<i>E. coli</i> ED1a (chromosome + plasmid)	B2	2.66×10^{-7}	7.18×10^{-7} (3.28×10^{-7} – 1.44×10^{-6})	6.31 (2.88–12.64)–22.10 (10.09–44.32)	This study	
<i>E. coli</i> clone D	B2	2.26×10^{-7b}	1.40×10^{-7} (2.97×10^{-8} – 2.77×10^{-7})	1.23 (0.26–2.43)–4.3 (0.91–8.52)	22	
<i>E. coli</i> 536	B2	1.21×10^{-6b}	NA ^d	NA	14	
<i>E. coli</i> REL606	A	2.16×10^{-7} (9.73×10^{-8} – 3.4×10^{-7}) ^b	NA	NA	28	

^aWe calculated the mutation rate per genome per year and per base per year of each of the genomic elements according to two modalities: the coalescent theory (by Bayesian inference using BEAST) and the theory of the independence of the evolution of the lineages.

^bData published previously.

^cThe values were computed using a range of estimated mutation rates between 0.89×10^{-10} mutations per base per generation published in Wielgoss et al. (28) and 3.12×10^{-10} mutations per base per generation published in Foster et al. (29).

^dNA, not available.

$G = \mu_{\text{est}}/(\mu_g \times 365)$ (see File S1 in the supplemental material), where μ_g is the mutation rate per base per generation of ED1a in the healthy human gut. μ_g is unknown; however, several generational mutation rates have been previously reported using genome-wide approaches under *in vitro* conditions (28–30). Even an estimation of the ED1a mutation rate per generation was reported in Foster et al. (29). However, the environment seems to play a major role in the occurrence of mutations by generations (31). It is indeed difficult to establish experimentally a mutation rate per generation for *E. coli* because the rate depends on the strain (29) and the culture medium (31). However, we observe that the rates of mutations by generation determined with different methods, different strains, and different nonstressful media have very close values (of the order of a maximum factor of 3.5 between the extreme values). We can therefore use the extreme estimates for our present study: 0.89×10^{-10} (28) and 3.12×10^{-10} (29) mutations per base per generation.

Assuming that the number of mutations per base per generation ranges from 0.89×10^{-10} to 3.12×10^{-10} , we can compute, using the chromosome, a range from 6.06 (95% CI, 2.88 to 12.65) to 21.24 (95% CI, 9.66 to 43.28) generations per day (considering lineages with genealogical links) occurring for the *E. coli* ED1a chromosome in healthy human gut. We also estimated the number of generations per day for clone D followed in a household for 3 years from 1.23 (95% CI, 0.26 to 2.43) to 4.3 (95% CI, 0.91 to 8.52) (Table 2).

Looking for traces of selection during *E. coli* ED1a evolution. We then decided to determine whether the genome sequences could reveal some signal of selection. To compute how the population behaved, we used a population genetics approach. The first step was to compute an estimate of diversity based on the genome sequences. For that purpose, we focused on point mutations on the chromosome as indels or mutations on the plasmid may have different mutation rates. Using only point mutations on the chromosome, the measure of diversity as estimated by the theta of Watterson (θ_w) (32), using the pegas package (33) in R (34), was 1.93 ± 1.13 , 1.54 ± 0.96 , and 1.15 ± 0.78 for sampling time points D0, D211, and D315, respectively. Using these later estimates, the genome length (L) and a range of mutation rates per base per generation from 0.89×10^{-10} to 3.12×10^{-10} (see above), as $\theta_w = 2N_eL\mu_g$, we estimated a range of effective population size (N_e) between 500 and 1,700. This means that the population studied has a standing genetic diversity that is equivalent to an idealized population of constant size N_e of ≈ 500 to 1,700 individuals. The observed diversity suggests therefore a small size.

We used Tajima's D to evaluate the deviation of the allele frequency distribution from that expected in a standard neutral evolution model (35). Taking into account point mutations on the chromosome, Tajima's D , using the pegas package (33) in R (34), was -0.84 ($P = 0.47$) at D0, -1.03 ($P = 0.35$) at D211, and -0.43 ($P = 0.74$) at D315. Thus, the evolution of these isolates did not deviate significantly from the neutral evolution model at each time point. Yet the values are all negative and could support selection of bottleneck events.

We further looked at synonymous and nonsynonymous mutations to infer if selection acting on nonsynonymous mutations could lead to a differential evolution between the two categories. Taking all different mutations found in the whole data set, the ratio of the rate of nonsynonymous to that of synonymous changes, K_a/K_s , was 1.3 (95% CI estimated by bootstrapping mutations, 0.51 to infinity). Though the power is quite limited due to the small number of mutations, the value of the ratio suggests that nonsynonymous mutations have accumulated as synonymous mutations in the different genomes and rejects an important contribution from selection. Furthermore, out of the four mutations that have been recovered in at least two isolates, one is synonymous, one is intergenic, and two are nonsynonymous, which suggests that among mutations that have survived drift to some "high" frequency, there is no excess of nonsynonymous mutations.

Among the four mutations that are not singletons, two mutations invaded the population over the year of the study. A synonymous mutation in the middle of gene *ybdA*, a putative autotransporter, was found in 50% (95% CI, 30% to 70%) of the isolates at D0 and fixed at D211. A mutation in the intergenic region between genes *ibpA* and *yidQ* is present in all isolates at D315, and none is present before. This mutation is in the promoter of *yidQ* that encodes an outer membrane protein whose function remains to be defined. While fixation is likely to reflect the contribution of selection, according to our estimates of the annual mutation rate and the mean diversity found at each time point, the average time to fix a neutral mutation should be, on average, 156 days. This corresponds to an average length of about 5.2 months for a neutral mutation to appear and invade the population, a time frame fully compatible with our observations. Moreover, given the annual mutation rate estimated at 3.32 mutations per genome per year, we can gauge that 1.66 neutral mutations that are destined to reach fixation should have occurred in the first 6 months of the study. Once again, this is a number that is close to the observation of one mutation fixing in the second half of the study. Hence, these rough estimations suggest that the fixation of the *yidQ* promoter mutation may result from genetic drift rather than from natural selection.

Signs of convergence, a signature of natural selection (12), were nevertheless observed in *rpoS*. Two different mutations on two different samples at D0 (I198S in 1D0 and V102M in 7D0) were recovered. These mutations had, however, no long-term selective advantage as they were just isolated from individual isolates from the early time points.

DISCUSSION

In the present work, we followed retrospectively the evolutionary history of an *E. coli* natural isolate: *E. coli* strain ED1a in its natural environment, the gut of a healthy human. The clone we studied exhibited the O81 type and belonged to a human-specific avirulent B2 subgroup, named subgroup VIII (ST452). It was the only *E. coli* we detected at the three time points we sampled, D0, D211, and D315, which implies full dominance of the *E. coli* niche over almost a full year of sampling.

The high dominance of the clone is consistent with several observations. First, B2 strains are recovered at high frequency among humans of industrialized countries (15, 36). Changes in hygiene and/or diet are suspected to contribute to that high prevalence. Over the last 4 decades, a little more than a 3-fold (36) increase in B2 frequency was observed in the gut of French individuals from whom our sampled individual was chosen. Second, when B2 strains are present within an individual, they are usually more dominant than other strains (37, 38). Third, B2 subgroup VIII clones, encompassing *E. coli* ED1a, were found to be dominant in 42% of hosts when found, which was higher than the value for other B2 group strains, which were found to be dominant in only 17% of hosts (21). This suggests an overall high adaptation of *E. coli* ED1a to the human gut.

Most changes observed during this year of evolution were SNPs (88%). For instance, we did not detect any phage integration or any gain or loss of plasmids. Based on SNPs, we could estimate an annual mutation rate, assuming a homogenous population evolving according to coalescent theory. The rate we estimated was 6.90×10^{-7} mutations per base per year on the chromosome and 7.18×10^{-7} mutations per base per year if the plasmid is taken into account. This is 3.2 times higher than the annual rate of 2.16×10^{-7} mutations per base per year found *in vitro* in the long-term evolution experiment (LTEE) of Richard Lenski (28). An annual rate 1.7 times higher, 1.21×10^{-6} mutations per base per year, was found in strains evolving in streptomycin-treated mice (14), and a rate five times lower, 1.4×10^{-7} mutations per base per year, was found using strains evolving within a household (22) (Table 2). The value we observed is hence intermediate between the estimates observed in the LTEE though the per-base per-generation mutation rate and number of generations per day may vary across conditions.

Assuming a constant mutation rate per generation that we estimated using a range

of published mutation rates, we can compute that *E. coli* is replicating between 6.06 (95% CI, 2.88 to 12.65) and 21.24 (95% CI, 9.66 to 43.28) times per day in the human gut (Table 2). The range of values found encompasses the 8 (39) and 18 (40) generations per day found previously in the gut of mice treated with streptomycin. The estimate of eight generations per day relied on the dilution of preinduced fluorescence and reflects an average over the population (39). The estimate of 18 generations per day relied on the quantification of ribosome content, a proxy for growth rate, using a 23S rRNA fluorescent oligonucleotide probe (40). These ranges of the numbers of divisions per day are compatible with *E. coli* keeping up with the mucus turnover rate (41, 42). We computed a lower number of generations per day, ranging from 1.23 (95% CI, 0.26 to 2.43) to 4.3 (95% CI, 0.91 to 8.52) in the case of *E. coli* clone D evolving in a household over 3 years (Table 2). Interestingly, in contrast to our clone that was dominant over almost a year, clone D evolved in the household but was neither dominant nor consistently recovered in any host. Transitions from one host to another require transitions of unspecified duration in secondary environments poor in nutrients and therefore supporting limited growth (43).

Experimental evolution over a year coupled to genomics, *in vitro* (6, 12) or *in vivo* (14), indicates that selection shapes genome evolution. Footprints of selection, such as overrepresentation of nonsynonymous mutations and traces of convergence, are overwhelming. We wondered whether we could find similar signals over almost 1 year of evolution in the human gut. To do so, we first looked at the number of synonymous and nonsynonymous mutations. Using the whole data set, we could not find a clear difference between the two types of mutations, suggesting that the mutations we observed were accumulated as neutral mutations. We then looked at convergence, and the only signal that emerged was from a couple of mutations found in *rpoS* at D0 which were not recovered later. Unfortunately, *rpoS* mutations have to be studied with caution as they emerge rapidly under laboratory conditions (44). Though our samples have experienced a minimum number of steps in the laboratory, it is still possible that laboratory evolution before DNA isolation could be responsible for these mutations. We have therefore no proof that selection is at play in the pattern of the mutations we observed.

Evolution without selection can be easily framed with population genetics. For instance, based on the number of polymorphic sites and mutation rate estimates, we computed the effective population size, N_e , as ranging from 500 to 1,700, the number of mutations appearing within a year in the population that has invaded the population in the long term, n_{fix} , as 3.32 (Table 2), and the mean time it takes for these mutations to reach fixation, T_{fix} , as 5.2 months (results independent of the per-generation mutation rate estimate). The observation of two fixations over almost a year, one starting from 50% and another one from a lower frequency, is fully compatible with neutral evolution. The low effective population size appeared nevertheless to be strikingly low, knowing that *E. coli* can be found in from 10^6 to 10^8 cells per gram of feces (45). It is, however, possible that a small fraction of the cells within the mucus may be founders of the rest of the population. Accordingly, previous work in streptomycin-treated mice has suggested that *E. coli* is not dividing in the lumen (46).

Several factors could explain the lack of a clear signal of selection over almost a year. First, the study of the LTEE has revealed that a high level of adaptation results in a lower fraction of beneficial mutations (6). However, even after 50,000 generations of evolution in a constant environment, more nonsynonymous than synonymous mutations were recovered. One alternative explanation is that the environment is not constant but fluctuating through time. As ED1a is well adapted, there may not be any clear way to improve over the average environment. Finally, the low effective population size may limit the opportunity for beneficial mutations to appear. Along these lines, it is worth noting that the two selective sweeps observed involved a single mutation and not a cohort of multiple mutations, as observed in experimental evolution (47). This confirms that if selection is at play, its strength, as measured by the product of population size and beneficial mutation rate, is very weak.

Our observations suggest that there was no clear sign of selection over almost a year of evolution in the human gut. This contrasts with the genomic evolution observed in experimental evolution. The data are indeed compatible with a small effective population size evolving neutrally even though the clone was dominant in its host over almost a year. Analyzing more longitudinal data of *E. coli* in the human gut will be required to test how general is this pattern of evolution.

MATERIALS AND METHODS

Sampling strategy. A single feces sample was collected from one healthy 44-year-old male living in Paris, France, over a period of almost 1 year, between October 2001 and September 2002, with three sampling points, at day 0 (D0), day 211 (D211) and day 315 (D315). The study was approved by the ethics evaluation committee of Institut National de la Santé et de la Recherche Médicale (INSERM). This person had no medical issues or diarrheic episodes and did not take any drug during the studied period and the year before. Moreover, this person has a Western lifestyle in terms of diet and hygiene. An aliquot of the fresh feces was spread on Drigalski agar plates that select for Gram-negative aerobic bacilli. The plates were incubated at 37°C over night (O/N), and the feces were discarded. Eight colonies of *E. coli* identified by a yellow appearance were isolated from each plate at each time point. These samples were then grown on liquid lysogeny broth (LB) O/N, stored at −80°C in glycerol, and thereafter called “isolate.” Finally, the 24 isolates obtained were identified by the numbers 1 to 8 and a suffix for the three time points of sampling, D0, D211, or D315. To confirm that the isolates belonged to the *E. coli* ED1a clone, we performed a PCR screening for the B2 phylogroup (48), followed by O81 and B2 subgroup VIII typing (21).

High-throughput sequencing. The glycerol stock of each of the 24 isolates was then plated on LB agar, and one colony was chosen to extract the total DNA by using a Wizard genomic DNA purification kit (Promega). The 24 genomes were then sequenced using an Illumina platform HiSeq instrument with an average of almost 22 million paired-end reads of 51 bp per read.

Read mapping and mutation identification. We used as a reference the genome of ED1a that was sequenced using Sanger technology, fully assembled, and annotated in a previous work (23). It is composed of a circular chromosome of 5,209,548 bp and a large conjugative plasmid of 119,594 bp. This sequenced clone is one of the clones sampled at day 0. We resequenced this clone with the same technology as that used with the other clones to identify errors in this reference using breseq, version 0.27.2 (49).

breseq uses Bowtie2 to map reads to a reference genome. It then identifies mutation evidence that can take the form of either read alignment (RA) evidence, corresponding to single nucleotide polymorphisms and short indels and missing coverage, and new junction evidence, corresponding to reads mapping to one part of the reference on one of its sides and to another part on the other side, indicating a possible rearrangement. The program then uses this evidence to make mutation predictions. The RAs are then transformed into predictions as SNPs or short indels when they are supported by at least 85% of the read. breseq can also identify a large deletion and chromosomal rearrangement when it is able to correlate a missing coverage in a region with new junction evidence on both sides of this region.

We then filtered the mutations identified by breseq. When filtering the outputs of breseq, we first looked at the predicted SNPs and short indels. We removed mutations that seemed to appear in each sample as those likely came from a sequencing error in the reference genome and were not informative on the dynamics of diversification during this experiment. We also removed mutations when they were too close to one another and discarded variations that were less than 51 bp apart. These clustered mutations are indeed usually caused by erroneously mapped reads, and previous analyses showed that they are typically found in prophagic regions. These mobile regions are repeated in the genome but do not have 100% identity, generating trouble in the mapping process as they are still close enough to one another to be erroneously mapped. Indeed, the phage-mediated exchange of DNA sequences among bacteria occurs with high frequency (50), resulting in constant modifications of specific regions of the genome. We also removed all mutations for which the frequency of the mutated reads was less than 0.95.

A total of 1,029 mutations with a mutant frequency higher than 0.95 were detected by breseq among the 24 isolates. On the chromosome, these correspond to 329 specific mutations. Among these, 192 were recovered in all strains, corresponding to sequencing errors in the Sanger-sequenced reference genome used for mapping the reads. A total of 122 other mutations were clustered in three genes, *ECED1_1442*, *ECED1_1710*, and *ECED1_2512*. All three genes are in prophagic regions that have multiple copies in the genome with some level of divergence. The mutations detected in these regions result most likely from assembly problems and/or the existence of partial gene conversion between the different copies. Furthermore, these mutations, while found in some genomes above the threshold read frequency of 0.95, are found at lower read frequencies in the other strains, confirming the idea of assembly issues in these regions. Finally, only 15 mutations were retained for analysis on the chromosome. On the plasmid, 700 specific mutations were recovered, 456 of which were sequencing errors in the reference. Out of the 244 remaining mutations, 242 were clustered in two regions, *pECED1a_0013/pECED1a_0014* and *insC/pECED1a_0084/pECED1a_0085*. Ultimately only two mutations were retained for analysis on the plasmid.

Prophage and plasmid detection in isolate sequences. Using the sequences obtained for the 24 isolates, we started by filtering them to eliminate all the reads matching the reference genome (5,209,548 bp) and plasmid (119,594 bp) (23). To this end, Bowtie2 was used with default parameters to align the reads to the reference and to keep only the unaligned reads (51). Among the 24 isolates, the number of

unaligned reads and their proportion are, respectively, around 1.5×10^6 and 7%. These unaligned reads were then assembled using SPAdes (52). To detect prophage sequences from these assembled unaligned reads, we used blastx with the PHASTER database (updated 3 August 2017) (53). Then, we assembled the remaining unaligned reads with plasmidSPAdes in order to capture plasmid sequences with greater accuracy (54). Prokka was used to annotate the plasmid sequences (25); we then used discoSnp++ to detect mutations between these potential plasmids (26).

In order to know if this hypothetical plasmid was merely the result of contamination or a bona fide genetic element present in the strain but undetected to this day, we designed a simple PCR assay with primers specific to the hypothetical plasmid matching an ORF coding a protein belonging to the *mobA-mobL* superfamily involved in bacterial conjugation (forward, 5'-GGTCTGCTCACCGCTTCT-3'; reverse, 5'-GCATGATTGCCGATGTGGCG-3').

Calculation of mutation rates and phylogenetic and selection analyses. In our study, we calculated the mutation rate per base per year of each of the genomic elements according to two models: the coalescent theory (by Bayesian inference) and the independence of the evolution of the lineages.

The unrooted phylogenetic tree was reconstructed by the maximum likelihood method with the packages ape (55) and phangorn (56) in R software (34). Regarding the selection indices, the calculation of diversity by the theta of Watterson (θ_w) and the statistical test Tajima's *D* were done using the pegas package (33) in the R software (34), and the calculation of K_a/K_e was generated taking into account the codon bias and the proportion of transitions/transversions. Calculations for diversity, estimation of the number of generations per day, mean time for fixation in generations, and mean time for fixation in days are based on population genetics assuming Kimura's neutral model, as described in File S1 in the supplemental material (57).

Data availability. The fastq files corresponding to each isolate and the genome reference used to analyze mutations and produce Fig. 1 and Tables 1 and 2 are accessible on the Dryad website (<https://doi.org/10.5061/dryad.v5374>).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.02377-17>.

SUPPLEMENTAL FILE 1, PDF file, 0.3 MB.

ACKNOWLEDGMENTS

This work was supported in part by grants from La Fondation pour la Recherche Médicale to Mohamed Ghalayini (thèse médico-scientifique, grant number FDM20150633803) and Erick Denamur (équipe FRM 2016, grant number DEQ20161136698) and from the European Research Council to Olivier Tenaillon under the European Union's Seventh Framework Program (FP7/2007–2013)/ERC grant 310944. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

We gratefully acknowledge Francois Cohen and Hervé Le Nagard for their valuable and indispensable help with the CATiBioMed cluster.

REFERENCES

- Bradford PA. 2001. Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. *Clin Microbiol Rev* 14:933–951. <https://doi.org/10.1128/CMR.14.4.933-951.2001>.
- Friedman ND, Temkin E, Carmeli Y. 2016. The negative impact of antibiotic resistance. *Clin Microbiol Infect* 22:416–422. <https://doi.org/10.1016/j.cmi.2015.12.002>.
- Ciorba V, Odone A, Veronesi L, Pasquarella C, Signorelli C. 2015. Antibiotic resistance as a major public health concern: epidemiology and economic impact. *Ann Ig* 27:562–579.
- van Buul LW, van der Steen JT, Veenhuizen RB, Achterberg WP, Schellevis FG, Essink RT, van Benthem BH, Natsch S, Hertogh CM. 2012. Antibiotic use and resistance in long term care facilities. *J Am Med Dir Assoc* 13:568.e1–e13. <https://doi.org/10.1016/j.jamda.2012.04.004>.
- Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, Taddei F. 2001. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* 291:2606–2608. <https://doi.org/10.1126/science.1056421>.
- Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, Wu GC, Wielgoss S, Cruveiller S, Médigue C, Schneider D, Lenski RE. 2016. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* 536:165–170. <https://doi.org/10.1038/nature18959>.
- Wiser MJ, Ribeck N, Lenski RE. 2013. Long-term dynamics of adaptation in asexual populations. *Science* 342:1364–1367. <https://doi.org/10.1126/science.1243357>.
- Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200:619–631. <https://doi.org/10.1534/genetics.115.176677>.
- Giraud A, Arous S, De Paepe M, Gaboriau-Routhiau V, Bambou J-C, Rakotobe S, Lindner AB, Taddei F, Cerf-Bensussan N. 2008. Dissecting the genetic components of adaptation of *Escherichia coli* to the mouse gut. *PLoS Genet* 4:e2. <https://doi.org/10.1371/journal.pgen.0040002>.
- Hindré T, Knibbe C, Beslon G, Schneider D. 2012. New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nat Rev Microbiol* 10:352–365. <https://doi.org/10.1038/nrmicro2750>.
- Barroso-Batista J, Sousa A, Lourenço M, Bergman M-L, Sobral D, Demengeot J, Xavier KB, Gordo I. 2014. The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. *PLoS Genet* 10:e1004182. <https://doi.org/10.1371/journal.pgen.1004182>.
- Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF,

- Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* 335:457–461. <https://doi.org/10.1126/science.1212986>.
13. Welling GW, Groen G, Tuinte JH, Koopman JP, Kennis HM. 1980. Biochemical effects on germ-free mice of association with several strains of anaerobic bacteria. *J Gen Microbiol* 117:57–63.
 14. Lescat M, Launay A, Ghalayini M, Magnan M, Glodt J, Pintard C, Dion S, Denamur E, Tenaillon O. 2016. Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an *Escherichia coli* natural isolate in the streptomycin-treated mouse gut. *Mol Ecol* 26:1802–1817. <https://doi.org/10.1111/mec.13851>.
 15. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8:207–217. <https://doi.org/10.1038/nrmicro2298>.
 16. Russo TA, Johnson JR. 2003. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes Infect* 5:449–456. [https://doi.org/10.1016/S1286-4579\(03\)00049-2](https://doi.org/10.1016/S1286-4579(03)00049-2).
 17. Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholary O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S. 2008. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC Genomics* 9:560. <https://doi.org/10.1186/1471-2164-9-560>.
 18. Johnson JR, Clermont O, Menard M, Kuskowski MA, Picard B, Denamur E. 2006. Experimental mouse lethality of *Escherichia coli* isolates, in relation to accessory traits, phylogenetic group, and ecological source. *J Infect Dis* 194:1141–1150. <https://doi.org/10.1086/507305>.
 19. Clermont O, Gordon D, Denamur E. 2015. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology* 161:980–988. <https://doi.org/10.1099/mic.0.000063>.
 20. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O. 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* 24:2373–2384. <https://doi.org/10.1093/molbev/msm172>.
 21. Clermont O, Lescat M, O'Brien CL, Gordon DM, Tenaillon O, Denamur E. 2008. Evidence for a human-specific *Escherichia coli* clone. *Environ Microbiol* 10:1000–1006. <https://doi.org/10.1111/j.1462-2920.2007.01520.x>.
 22. Reeves PR, Liu B, Zhou Z, Li D, Guo D, Ren Y, Clabots C, Lan R, Johnson JR, Wang L. 2011. Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. *PLoS One* 6:e26907. <https://doi.org/10.1371/journal.pone.0026907>.
 23. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiappello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguénec C, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tournet J, Vacherie B, Vallenet D, Médigue C, Rocha EPC, Denamur E. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5:e1000344. <https://doi.org/10.1371/journal.pgen.1000344>.
 24. Mukherjee S, Huntemann M, Ivanova N, Kyrpides NC, Pati A. 2015. Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Stand Genomic Sci* 10:18. <https://doi.org/10.1186/1944-3277-10-18>.
 25. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 26. Uricaru R, Rizk G, Lacroix V, Quillery E, Plantard O, Chikhi R, Lemaître C, Peterlongo P. 2015. Reference-free detection of isolated SNPs. *Nucleic Acids Res* 43:e11. <https://doi.org/10.1093/nar/gku1187>.
 27. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214. <https://doi.org/10.1186/1471-2148-7-214>.
 28. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chané-Woon-Ming B, Médigue C, Lenski RE, Schneider D. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A* 110:222–227. <https://doi.org/10.1073/pnas.1219574110>.
 29. Foster PL, Lee H, Popodi E, Townes JP, Tang H. 2015. Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proc Natl Acad Sci U S A* 112:E5990–E5999. <https://doi.org/10.1073/pnas.1512136112>.
 30. Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci U S A* 109:E2774–E2783. <https://doi.org/10.1073/pnas.1210309109>.
 31. Maharjan RP, Ferenci T. 2017. A shifting mutational landscape in 6 nutritional states: stress-induced mutagenesis as a series of distinct stress input-mutation output relationships. *PLoS Biol* 15:e2001477. <https://doi.org/10.1371/journal.pbio.2001477>.
 32. Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7:256–276. [https://doi.org/10.1016/0040-5809\(75\)90020-9](https://doi.org/10.1016/0040-5809(75)90020-9).
 33. Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419–420. <https://doi.org/10.1093/bioinformatics/btp696>.
 34. R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
 35. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
 36. Massot M, Daubie AS, Clermont O, Jauréguy F, Couffignal C, Dahbi G, Mora A, Blanco J, Branger C, Mentré F, Eddi A, Picard B, Denamur E, The Coliville Group. 2016. Phylogenetic, virulence and antibiotic resistance characteristics of commensal strain populations of *Escherichia coli* from community subjects in the Paris area in 2010 and evolution over 30 years. *Microbiology* 162:642–650. <https://doi.org/10.1099/mic.0.000242>.
 37. Blyton MDJ, Cornall SJ, Kennedy K, Colligon P, Gordon DM. 2014. Sex-dependent competitive dominance of phylogenetic group B2 *Escherichia coli* strains within human hosts. *Environ Microbiol Rep* 6:605–610. <https://doi.org/10.1111/1758-2229.12168>.
 38. Smati M, Clermont O, Le Gal F, Schichmanoff O, Jauréguy F, Eddi A, Denamur E, Picard B, Coliville Group. 2013. Real-time PCR for quantitative analysis of human commensal *Escherichia coli* populations reveals a high frequency of subdominant phylogroups. *Appl Environ Microbiol* 79:5005–5012. <https://doi.org/10.1128/AEM.01423-13>.
 39. Myhrvold C, Kotula JW, Hicks WM, Conway NJ, Silver PA. 2015. A distributed cell division counter reveals growth dynamics in the gut microbiota. *Nat Commun* 6:10039. <https://doi.org/10.1038/ncomms10039>.
 40. Rang CU, Licht TR, Midtvedt T, Conway PL, Chao L, Krogfelt KA, Cohen PS, Molin S. 1999. Estimation of growth rates of *Escherichia coli* BJ4 in streptomycin-treated and previously germfree mice by in situ rRNA hybridization. *Clin Diagn Lab Immunol* 6:434–436.
 41. Conway T, Cohen PS. 2015. Commensal and pathogenic *Escherichia coli* metabolism in the gut. *Microbiol Spectr* 3:3. <https://doi.org/10.1128/microbiolspec.MBP-0006-2014>.
 42. Johansson MEV. 2012. Fast renewal of the distal colonic mucus layers by the surface goblet cells as measured by in vivo labeling of mucin glycoproteins. *PLoS One* 7:e41009. <https://doi.org/10.1371/journal.pone.0041009>.
 43. Ihssen J, Egli T. 2005. Global physiological analysis of carbon- and energy-limited growing *Escherichia coli* confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. *Environ Microbiol* 7:1568–1581. <https://doi.org/10.1111/j.1462-2920.2005.00846.x>.
 44. Bleibtreu A, Clermont O, Darlu P, Glodt J, Branger C, Picard B, Denamur E. 2014. The *rpoS* gene is predominantly inactivated during laboratory storage and undergoes source-sink evolution in *Escherichia coli* species. *J Bacteriol* 196:4276–4284. <https://doi.org/10.1128/JB.01972-14>.
 45. Ervin JS, Russell TL, Layton BA, Yamahara KM, Wang D, Sassoubre LM, Cao Y, Kelly CA, Sivaganesan M, Boehm AB, Holden PA, Weisberg SB, Shanks OC. 2013. Characterization of fecal concentrations in human and other animal sources by physical, culture-based, and quantitative real-time PCR methods. *Water Res* 47:6873–6882. <https://doi.org/10.1016/j.watres.2013.02.060>.
 46. Poulsen LK, Lan F, Kristensen CS, Hobolth P, Molin S, Krogfelt KA. 1994. Spatial distribution of *Escherichia coli* in the mouse large intestine inferred from rRNA in situ hybridization. *Infect Immun* 62:5191–5194.
 47. Lang GI, Desai MM. 2014. The spectrum of adaptive mutations in experimental evolution. *Genomics* 104:412–416. <https://doi.org/10.1016/j.ygeno.2014.09.011>.
 48. Clermont O, Christenson JK, Denamur E, Gordon DM. 2013. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5:58–65. <https://doi.org/10.1111/1758-2229.12019>.
 49. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data us-

- ing breseq. *Methods Mol Biol* 1151:165–188. https://doi.org/10.1007/978-1-4939-0554-6_12.
50. KENZAKA T, TANI K, SAKOTANI A, YAMAGUCHI N, NASU M. 2007. High-frequency phage-mediated gene transfer among *Escherichia coli* cells, determined at the single-cell level. *Appl Environ Microbiol* 73:3291–3299. <https://doi.org/10.1128/AEM.02890-06>.
51. LANGMEAD B, SALZBERG SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
52. BANKEVICH A, NURK S, ANTIPOV D, GUREVICH AA, DVORKIN M, KULIKOV AS, LESIN VM, NIKOLENKO SI, PHAM S, PRJIBELSKI AD, PYSHKIN AV, SIROTKIN AV, VYAHHI N, TESLER G, ALEKSEYEV MA, PEVZNER PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
53. ARNDT D, GRANT JR, MARCU A, SAJED T, PON A, LIANG Y, WISHART DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkw387>.
54. ANTIPOV D, HARTWICK N, SHEN M, RAIKO M, LAPIDUS A, PEVZNER PA. 2016. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32:3380–3387. <https://doi.org/10.1093/bioinformatics/btv688>.
55. POPESCU A-A, HUBER KT, PARADIS E. 2012. ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* 28:1536–1537. <https://doi.org/10.1093/bioinformatics/bts184>.
56. SCHLIEP KP. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27:592–593. <https://doi.org/10.1093/bioinformatics/btq706>.
57. NIELSEN R, SLATKIN M. 2013. An introduction to population genetics: theory and applications. Sinauer Associates, Sunderland, Mass.